AD-A091 607 FLORIDA STATE UNIV TALLAHASSEE DEPT OF STATISTICS F/G 12/1
MULTIPARAMETER HYPOTHESIS TESTING AND ACCEPTANCE SAMPLING.(U)
JUL 80 R L BERGER DAAG29-79-C-0158
UNCLASSIFIED FSU-STATISTICS-M555 ARO-16713.12-M NL

END
DATE
FILMED
12 80
DTIC

UNCLASSIFIED 18 ARO 1

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

LEVEL II

11

AD A091607

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: 16713.12-M
2. GOVT ACCESSION NO.: AD-A091 607
3. RECIPIENT'S CATALOG NUMBER:
4. TITLE (and Subtitle): Multiparameter Hypothesis Testing and Acceptance Sampling
5. TYPE OF REPORT & PERIOD COVERED: Technical rept.
6. PERFORMING ORG. REPORT NUMBER:
7. AUTHOR(s): Roger L. Berger
8. CONTRACT OR GRANT NUMBER(s): DAAG29-79-C-0158
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Florida State University, Tallahassee, FL 32306
10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS:
11. CONTROLLING OFFICE NAME AND ADDRESS: U. S. Army Research Office, Post Office Box 12211, Research Triangle Park, NC 27709
12. REPORT DATE: Jul 80
13. NUMBER OF PAGES: 9
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): 12 12
15. SECURITY CLASS. (of this report): Unclassified
15a. DECLASSIFICATION/DOWNGRADING SCHEDULE:
16. DISTRIBUTION STATEMENT (of this Report): Approved for public release; distribution unlimited. 14 FSU-STATISTICS-M555, TR-D48

DTIC ELECTE NOV 1 2 1980 S D B

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report): NA
18. SUPPLEMENTARY NOTES: The view, opinions, and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.
19. KEY WORDS (Continue on reverse side if necessary and identify by block number):
   consumer's risk
   multiple inference
   uniformly most powerful
20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The quality of a product might be determined by several parameters, each of which must meet certain standards before the product is acceptable. In this paper, a method of determining whether all the parameters meet their respective standards is proposed. The method consists of testing each parameter individually and deciding that the product is acceptable only if each parameter passes its test. This simple method has some optimal properties including attaining exactly a pre-specified consumer's risk and uniformly minimizing the producer's risk. These results are obtained from more general hypothesis testing results concerning null hypothesis consisting of the unions of sets.

DDC FILE COPY

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

400277

80 11 03 102

MULTIPARAMETER HYPOTHESIS TESTING AND
ACCEPTANCE SAMPLING

by

Roger L. Berger
The Florida State University

FSU Statistics Report M555
USARO Technical Report D48

The Florida State University
Department of Statistics
Tallahassee, Florida 32306

July, 1980

[1]This research was supported by the U.S. Army Research Office Grant No. DAAG29-79-C-0158.

Roger L. Berger is Assistant Professor, Department of Statistics, The Florida State University, Tallahassee, Florida 32306.

ABSTRACT

The quality of a product might be determined by several parameters, each of which must meet certain standards before the product is acceptable. In this paper, a method of determining whether all the parameters meet their respective standards is proposed. The method consists of testing each parameter individually and deciding that the product is acceptable only if each parameter passes its test. This simple method has some optimal properties including attaining exactly a prespecified consumer's risk and uniformly minimizing the producer's risk. These results are obtained from more general hypothesis testing results concerning null hypotheses consisting of the unions of sets.

Key words: consumer's risk, multiple inference, uniformly most powerful.

1. INTRODUCTION

In many situations, the quality of a product is determined by several parameters. The product is of acceptable quality to the consumer only if each of the parameters meets certain standards. For example, an upholstery fabric must meet standards for strength, colorfastness, and fire resistance. Based on some measurements on the product, the consumer must decide whether the product is acceptable, i.e., all of the parameters meet the standards, or unacceptable, i.e., one or more of the parameters do not meet the standards. In making this decision the consumer wishes to use a rule which controls the consumer's risk at a small level.

If there is only one parameter and only one kind of measurement, then a standard quality controls text such as Burr (1976) or Duncan (1974) gives methods for making this decision. Different methods are given depending on whether the parameter is a mean, variance or proportion of defectives and on whether the measurements are counts of defective units (sampling by attributes) or measurements on a continuous variable (sampling by measurements). But no text that the author has found deals with the situation in which there are multiple parameters of interest.

This problem will be formulated as a hypothisis testing problem in which the null hypotheses states that one or more of the parameters do not meet their standards and the alternative hypothesis states that all of the parameters do meet their standards. Then the probability of a Type I error will be the consumer's risk. Thus an α-level test will be one which controls the consumer's risk at less than or equal to α.

The test proposed herein is so simple it must not be new. But the author has not been able to find it described in hypotheses testing or quality control literature. The test is the following. A hypothesis test is done on each parameter individually at level α. The overall test rejects the null hypothesis and decides that all of the parameters meet their standards if and only if each individual test decides that the individual parameter meets its standard.

This test has several interesting properties. First, the individual tests are performed at level α and yet the overall test has level α. Usually when doing simultaneous inference about many parameters (see, e.g., Miller (1966)) inferences about individual parameters must be done with an error rate of less than α to achieve an overall error rate of α. This, for example is the basis of the Bonferroni method of simultaneous confidence intervals. Second, under very mild conditions, the level of this test is exactly α. So the test is not being too conservative by requiring each of the individual tests to decide that the individual parameter meets its standard. Third, under more restrictive conditions a result of Lehmann (1952) can be used to prove that this test is uniformly most powerful in a reasonable class of tests. In terms of risks this says that this test uniformly minimizes the producer's risk. These properties indicate that not only is the test extremely easy to implement, since it deals with only one parameter at a time, but it also seems to be a reasonably good test.

2. Basic Results

Let $X = (X_1, \cdots, X_n)$ be a random vector of observations whose distribution is determined by a vector parameter $\underset{\sim}{\theta} = (\theta_1, \cdots, \theta_\ell)$. Let Θ denote the parameter space. Let Θ_i, $i = 1, \cdots, K$, be subsets of Θ. Let $\Theta_0 = \bigcup_{i=1}^{K} \Theta_i$. Let A' denote the complement of the set A. Note that $\Theta_0' = \bigcap_{i=1}^{K} \Theta_i'$. The problem to be considered is that of testing H_0: $\theta \in \Theta_0$ vs H_1: $\theta \in \Theta_0'$. In the example in the introduction, Θ_i is the hypothesis that θ_i does not meet its standard.

If θ_i must be greater than c_i in order to meet its standard then $\Theta_i=\{\underset{\sim}{\theta}:\theta_i\le c_i\}$. If θ_i must be between c_i and d_i to meet its standard, then $\Theta_i=\{\underset{\sim}{\theta}:\theta_i\ge d_i \text{ or } \theta_i\le c_i\}$. With this formulation, H_0 is the hypothesis that at least one parameter does not meet its standard and H_1 is the hypothesis that every parameter meets its standard. Note that K, the number of subsets, may be less than ℓ, the number of parameters. This will be the case if some of the parameters are nuisance parameters and do not have standards associated with them.

Let α, $0\le\alpha\le1$ be fixed. For $i=1,\cdots,K$, let $\psi_i(\underset{\sim}{x})$ be an α-level test of $H_{0i}:\underset{\sim}{\theta}\epsilon\Theta_i$ vs $H_{1i}:\theta\epsilon\Theta'_i$, i.e., $E_{\underset{\sim}{\theta}}\psi_i(\underset{\sim}{X})\le\alpha$ for all $\underset{\sim}{\theta}\epsilon\Theta_i$. Let Ψ be the test of H_0 vs H_1 which rejects H_0 if and only if every ψ_i rejects H_{0i}.

Other authors such as Birnbaum (1954), Birnbaum, (1955), Lehmann (1955) and Spjotvoll (1972) have consider testing hypotheses H_0 and H_1. But in all these papers, except the one result of Lehmann to be discussed in Section 3, the null hypothesis is of the form H_1.

Tsutakawa and Hewett (1978) propose the test Ψ for a problem comparing regression lines. Theorems 1 and 2, which follow, are board generalizations of a result they prove about a test based on a bivariate t distribution.

Wilkinson (1951) has proposed a test like Ψ but in a very different situation. Wilkinson assumes that the individual tests are α-level tests for all of H_0, not just H_{0i}. Wilkinson also assumes the individual tests are independent which typically will not be the case in the problems considered herein.

The facts that Ψ is always an α-level test and under mild conditions has size exactly equal to α are presented in the following two theorems.

Theorem 1. Ψ is an α-level test of H_0 vs H_1, i.e., $E_{\underset{\sim}{\theta}}\Psi(\underset{\sim}{X})\le\alpha$ for all $\underset{\sim}{\theta}\epsilon\Theta_0$.

<u>Proof</u>: Let R_i be the event that ψ_i rejects H_{0i}. Then $R=\bigcap_{i=1}^{K}R_i$ is the event that Ψ rejects H_0. Fix $\underset{\sim}{\theta}\in\Theta_0$. Then $\underset{\sim}{\theta}\in\Theta_i$ for some i and

$E_{\underset{\sim}{\theta}}\,\Psi\,(\underset{\sim}{X})=P_{\underset{\sim}{\theta}}\,(R)\le P_{\underset{\sim}{\theta}}\,(R_i)=E_{\underset{\sim}{\theta}}\,\psi_i(\underset{\sim}{X})\le\alpha$ since $\underset{\sim}{\theta}\in\Theta_i$ and ψ_i is an α-level test of

H_{0i}. ||

Theorem 2: Suppose $\Theta_i = \{\underset{\sim}{\theta}: \theta_i \le c_i\}$, $i=1,\cdots, K$. Suppose the power of ψ_i depends only on θ_i and $\theta_{K+1},\cdots, \theta_\ell$. Suppose $E_{\underset{\sim}{\theta}}\psi_i(\underset{\sim}{X})=\alpha$ if $\theta_i=c_i$. Suppose there are upper bounds b_i (possibly infinite) such that for any fixed values of $\theta_{K+1},\cdots, \theta_\ell$, $\lim_{\theta_i \to b_i} E_{\underset{\sim}{\theta}} \psi_i(\underset{\sim}{X})=1$. Then Ψ has size exactly α, i.e., $\sup_{\underset{\sim}{\theta}\in\Theta_0} E_{\underset{\sim}{\theta}} \Psi(\underset{\sim}{X}) = \alpha$.

<u>Proof:</u> Let R_i and R be defined as in the proof of Theorem 1. Let $\underset{\sim}{\theta}_i = (\theta_{1i},\cdots, \theta_{\ell i})$, $i=1,2,\cdots$ be a sequence of parameter points satisfying $\theta_{1i} = c_1$ for all i, $\theta_{ji} \to b_j$, $j=2, \cdots, K$, and $\theta_{(K+1)i}, \cdots, \theta_{\ell i}$ are fixed for all i. Then $\underset{\sim}{\theta}_i \in \Theta_1$ for all i and $P_{\underset{\sim}{\theta}}(R_1') = 1-E_{\underset{\sim}{\theta}_i} \psi_1(\underset{\sim}{X}) = 1-\alpha$ for all i. Also, for $j=2, \cdots, K$, $\lim_{i\to\infty} P_{\underset{\sim}{\theta}_i}(R_j') = 1- \lim_{i\to\infty} P_{\underset{\sim}{\theta}_i}(R_j) = 1-1=0$.
Therefore,

$$
\begin{aligned}
\sup_{\underset{\sim}{\theta}\in\Theta_0} E_{\underset{\sim}{\theta}} \Psi(\underset{\sim}{X}) &\ge \lim_{i\to\infty} E_{\underset{\sim}{\theta}_i} \Psi(\underset{\sim}{X}) \\
&= \lim_{i\to\infty} P_{\underset{\sim}{\theta}_i}\Big(\bigcap_{j=1}^{K} R_j\Big) \\
&= \lim_{i\to\infty} \Big(1 - P_{\underset{\sim}{\theta}_i}\Big(\bigcup_{j=1}^{K} R_j'\Big)\Big) \\
&\ge 1 - \lim_{i\to\infty} \sum_{j=1}^{K} P_{\underset{\sim}{\theta}_i}(R_j') \\
&= 1 - \lim_{i\to\infty} \Big((1-\alpha) + \sum_{j=2}^{K} P_{\underset{\sim}{\theta}_i}(R_j')\Big) \\
&= 1 - (1-\alpha) - 0 = \alpha.
\end{aligned}
$$

From Theorem 1, $\sup_{\underset{\sim}{\theta}\in\Theta_0} E_{\underset{\sim}{\theta}} \Psi(\underset{\sim}{X}) \le \alpha$.

Thus the size of Ψ is exactly α. ||

The proof of Theorem 2 shows that for the test Ψ the maximum of the consumer's risk, the maximum of the probability of a Type I error, occurs when one parameter just fails the standard, $\theta_1 = c_1$, and all the other parameters are very good (large). This is exactly what would be expected. If many of the parameters do not meet their standards, it should be easy to decide that the product is unacceptable.

But if all the parameters but one are very good and the one exception nearly meets the standard, it will be most difficult to decide that the product is unacceptable.

The conditions of Theorem 2 will hold if θ_i is a normal mean and ψ_i is a one tailed t-test or if θ_i is the proportion of non-defective items and ψ_i is a one tailed binomial test. An important case in which the conditions of Theorem 2 will not hold is the case in which θ_i is a normal mean, the standard is $c_i < \theta_i < d_i$ and c_i and d_i are finite numbers. No α-level test will have a power approaching one on this set of alternatives.

3. <u>An optimality Result.</u>

Lehmann (1952) considered multiparameter hypothesis tests. He was primarily concerned with testing the null hypothesis H_1 vs the alternative H_0. But he proved one result, Theorem 4.2, for the situation in which H_0 is the null hypothesis. Although Lehmann did not speak in terms of combining individual tests to get an overall test, his Theorem 4.2 says that under certain conditions the test Ψ we have proposed is uniformly most powerful in a certain class. The remainder of this section is a review of Lehmann's result in terms of our notation.

A subset of A of R^k is called monotone if $\underset{\sim}{x} \in A$ and $y_i \geq x_i$, $i = 1, \cdots, K$, implies $\underset{\sim}{y} \in A$. Suppose $K = \ell$. The parameter space is the finite or infinite open rectangle $a_i < \theta_i < b_i$, $i = 1, \cdots, K$. Let $Y_1, \cdots Y_k$ denote K statistics. Suppose $p_{\underset{\sim}{\theta}}(\underset{\sim}{y})$ is the joint density of the Y's and the positive sample space of the Y's is the open rectangle $u_i < y_i < v_i$ independent of the θ's. Suppose the marginal distribution of Y_i depends only on θ_i and Y_i converges in probability to v_i as $\theta_i \to b_i$. Suppose the joint distribution of the Y's has the property that if $\underset{\sim}{\theta}$ and $\underset{\sim}{\theta}'$ are two parameter points with $\theta_i \leq \theta_i'$, $i = 1, \cdots K$, then

$P_{\underset{\sim}{\theta}}(Y \in A) \le P_{\underset{\sim}{\theta}'}(\underset{\sim}{Y} \in A)$ for every monotone set A. Let $\Theta_i = \{\underset{\sim}{\theta} : \theta_i \le c_i\}$. Let ψ_i be the test which rejects H_{0i} if $Y_i > w_i$ where w_i is chosen so that $P_{\underset{\sim}{\theta}}(Y_i > w_i) = \alpha$ if $\theta_i = c_i$. Under these assumptions the following result can be proven.

Theorem 3. (Lehmann's Theorem 4.2).

Under the above assumptions, the test Ψ which rejects $H_0 : \underset{\sim}{\theta} \in \Theta_0$ if and only if each ψ_i rejects $H_{0i} : \underset{\sim}{\theta} \in \Theta_i$ is uniformly most powerful among all non-randomized tests which have monotone rejection regions.

Although Lehmann did not discuss nuisance parameters, the results of Theorem 3 will continue to hold if the assumptions hold for each fixed set of values for $\theta_{K+1}, \cdots, \theta_\ell$.

Lehmann gives some techniques for verifying the monotonicity property assumed about the joint distribution of $Y_1, \cdots, Y_k$. It will hold, for example, if the Y's are independent t-statistics and the θ_i's are normal means.

4. AN EXAMPLE

An example of specifications given in terms of many parameters may be found in the textile industry. Table 1 lists specifications for upholstery fabric from the American Society for Testing and Materials. The specifications give standards for nine parameters related to strength, dimensional stability, colorfastness and flammability.

The first four parameters might be assumed to be normal means; mean breaking strength, mean tear strength, etc. The first three standards say the mean must be greater than some value. The dimensional change standard gives an upper and lower bound for the mean. These four hypotheses might be tested using t-tests. An upper bound on the variance of the dimensional change measurements will have to be assumed in order to construct an α-level test based on a t-statistics because

of the hypothesis' two sided form.

The last five parameters might be measured by binomial variables, each variable counting the number of units in a sample which pass the corresponding test. Each parameter would then be the proportion of units in the population (a particular manufacturer's output) which achieve one of the standards. The usual binomial test could be used to test each of these five, one-sided hypotheses.

It would be very difficult to posit a realistic multivariate model for these nine variables. Some are discrete and some are continuous. Some are likely to be correlated. Yet it is relatively easy to construct an α-level test for each parameter individually.

These tests can be combined into the overall Ψ to test the hypothesis H_0 with a consumer's risk of α. Theorems 2 and 3 indicate that, in certain situations, the resulting acceptance plan is fairly efficient in terms of both producer's and consumer's risks.

5. CONCLUSIONS

Acceptance sampling procedures for individual parameters are well known. This paper proposes a way of combining these procedures in the situation in which the quality of a product is measured by standards on several parameters. Not only is the method easy to implement but it controls the consumer's risk at exactly a pre-assigned level in typical situations when the standards are one-sided (either upper or lower bounds). Under slightly more restrictive conditions this method also uniformly minimizes the producer's risk. This method can be used in hypothesis testing problems other than acceptance sampling if the null hypothesis is a union of sets.

TABLE 1

Standard Specification for Woven Upholstery Fabric
Plain, Tufted or Flocked

Test	Minimum Standard
Breaking Strength	50 pounds
Tongue Tear Strength	6 pounds
Surface Abrasion (heavy duty)	15,000 cycles
Dimensional Change	5% shrinkage, 2% gain
Colorfastness to:	
Water	class 4
Crocking	
Dry	class 4
Wet	class 3
Light-40 AATCCF Fading Units	class 4
Flammability	Pass

Source: 1978 Annual Book of ASTM Standards, American Society for Testing and Materials, Philadelphia, Part 32, page 717.

REFERENCES

Birnbaum, Allen (1954) "Combining Independent Tests of Significance", Journal of the American Statistical Association, 49, 559-574.

________ (1955) "Characterizations of Complete Classes of Some Multiparameter Hypotheses, with Applications to Likelihood Ratio Tests", Annals of Mathematical Statistics, 26, 21-36.

Burr, Irving W. (1976) Statistical Quality Control Methods, Marcel Dekker, Inc., New York.

Duncan, Acheson J. (1974) Quality Control and Industrial Statistics, 4th edition, Richard D. Irvin, Inc., Homewood, IL.

Miller, Rupert G. (1966) Simultaneous Statistical Inference, McGraw-Hill, Inc., New York.

Lehmann, E.L. (1955) "Testing Multiparameter Hypotheses", Annals of Mathematical Statistics, 23, 541-552.

Spjotvoll, Emil (1972) "On the Optimality of Some Multiple Comparison Procedures," Annals of Mathematical Statistics, 43, 398-411.

Tsutakawa, R.K. and Hewett, J.E. (1978) "Comparison of Two Regression Lines Over a Finite Interval," Biometrics, 34, 391-398.

Wilkinson, B. (1951) "A Statistical Consideration in Psychological Research", Psychological Bulletin, 48, 156-157.

DAT
FILM
2-